

A modified logarithmic Sobolev inequality for the Hamming cube and some applications

Alex Samorodnitsky

Abstract

The logarithmic Sobolev inequality [9] for the Hamming cube $\{0, 1\}^n$ states that for any real-valued function f on the cube holds

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq 2 \cdot \text{Ent}(f^2)$$

We show that the constant $C = 2$ at the right hand side of this inequality can be replaced by a function $C(\rho)$ depending on $\rho = \frac{1}{n} \frac{\text{Ent}(f^2)}{\mathbb{E}f^2}$. The function $C(\cdot)$ is an increasing convex function taking $[0, \log 2]$ to $[2, 2/\log 2]$.

We present some applications of this modified inequality. In particular, it is used to obtain a discrete version of the Faber-Krahn inequality for small subsets of the Hamming cube, answering a question of Friedman and Tillich [8]. We introduce, following [8], the notion of a fractional edge-boundary size of a subset of $\{0, 1\}^n$, and show Hamming balls of radius at most $n/2 - \tilde{\Omega}(n^{3/4})$ to be sets with (asymptotically) the smallest fractional edge-boundary for their size.

1 Introduction

1.1 Isoperimetric problems on the Hamming cube

This paper deals with *discrete isoperimetric inequalities* on graphs. Let a graph $G = (V, E)$ be given, and let $A \subseteq V$ be a set of vertices in G . An isoperimetric inequality addresses the question of how small the boundary ∂A of A can be, given the cardinality of A , by lower bounding the size of ∂A by an appropriate function of $|A|$. There are various ways to define and measure the boundary of the set, two salient examples being the *vertex boundary* of A , consisting of the vertices of A which have neighbors outside A , and the *edge boundary* of A , which is the set of edges crossing from A to its complement. In both these cases, the size of the boundary is the cardinality of the corresponding set of vertices (or edges). However, one can also think of examples of a somewhat different nature, some of which will be considered below.

In this paper we deal with a specific graph - the Hamming cube $\{0, 1\}^n$. This is a graph with 2^n vertices indexed by boolean strings of length n . Two vertices are connected by an edge if they differ only in one coordinate. The metric defined by this graph is called the Hamming distance. In other words, two vertices x and y are at distance d if they differ in d coordinates.

Let us define two important families of subsets of $\{0,1\}^n$. A *Hamming ball* is a ball in the Hamming metric. A *subcube* is a subset of the vertices obtained by fixing the value in some of the coordinates. The number of fixed coordinates is called the co-dimension of the subcube. It turns out [11, 10, 12] that a Hamming ball has the smallest vertex boundary for its size, and a subcube has the smallest edge boundary.

We will consider several versions of the edge-isoperimetric inequality for the cube. Let $|\partial A|$ be the cardinality of the edge boundary of A , normalized, for convenience, by 2^{n-1} . The standard version of the inequality [10, 12] states that for any subset $A \subseteq \{0,1\}^n$ we have ¹

$$|\partial A| \geq \frac{2}{\log 2} \cdot \frac{|A|}{2^n} \log \frac{2^n}{|A|}. \quad (1)$$

This is tight if A is a subcube of an arbitrary co-dimension $0 \leq t \leq n$.

A *logarithmic Sobolev inequality* [9] establishes a relation between two (appropriately defined) notions: the variation of a function and its entropy. For a function f on the Hamming cube $\{0,1\}^n$ endowed with the uniform measure, this translates to

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq 2 \cdot \text{Ent}(f^2) = 2 \cdot (\mathbb{E} f^2 \log f^2 - \mathbb{E} f^2 \log \mathbb{E} f^2) \quad (2)$$

It is useful to view (2) as an isoperimetric inequality. In particular ([7]), it implies a functional form of the edge-isoperimetric inequality (1). For a non-zero function $f : \{0,1\}^n \rightarrow \mathbb{R}$ holds

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq 2 \cdot \mathbb{E} f^2 \log \frac{\mathbb{E} f^2}{\mathbb{E}^2 |f|} \quad (3)$$

Choosing f in (3) to be the characteristic function of a subset A of the cube, we recover the edge-isoperimetric inequality (1) with a somewhat worse constant on the right hand side, replacing $2/\log 2$ with 2. On the other hand, there are examples ([7]) of real-valued functions f for which the constant 2 in (3) is tight.

Let us now briefly sketch the results in this paper, before discussing them in fuller detail in subsection 1.4 below. We will show that the constant $C = 2$ at the right hand side of the logarithmic Sobolev inequality (2) can be replaced by a function $C(\rho)$ depending on $\rho = \frac{1}{n} \frac{\text{Ent}(f^2)}{\mathbb{E} f^2}$. The function $C(\rho)$ will be given explicitly. It is a convex function which increases from 2 to $2/\log 2$ as ρ goes from 0 to $\log 2$. We will also observe that $C(\rho)$ gives the correct dependence on ρ , describing functions on $\{0,1\}^n$ for which the modified inequality is tight.

This will imply a corresponding modification of the functional isoperimetric inequality (3). Here, as well as in (2), it will be possible to replace the constant $C = 2$ by the function $C(\rho)$.

The modified version of (3) is used to derive a discrete analogue of the classical Faber-Krahn inequality in \mathbb{R}^n for small subsets of the Hamming cube $\{0,1\}^n$, answering² a question

¹We use natural logarithms throughout the paper.

²Up to an error which becomes negligible as the dimension of the cube grows.

from [8]. We will introduce, following [8], the notion of a fractional edge-boundary size of a subset of $\{0, 1\}^n$, and show Hamming balls of radius at most $n/2 - \tilde{\Omega}(n^{3/4})$ to be sets with (asymptotically) the smallest fractional edge-boundary for their size.

The question in [8] is a part of an approach to obtain upper bounds on the cardinality of binary error-correcting codes. We will now take a brief detour to the theory of error-correcting codes in order to provide a natural framework in which this question can be discussed.

1.2 Bounds on binary error correcting codes

A binary error-correcting code of length n and minimal distance d is a subset C of the boolean cube $\{0, 1\}^n$ such that the distance between any two distinct points in C is at least d . In other words, the points in C can be taken as centers in a disjoint packing of Hamming balls of radius $\lceil \frac{d-1}{2} \rceil$ into $\{0, 1\}^n$.

The question of the maximal possible cardinality $A(n, d)$ of such a packing is one of the central questions of coding theory. The best known upper bounds on $A(n, d)$ were obtained in [17] following Delsarte's linear programming approach [6]. The analysis in [17] uses theory of orthogonal polynomials and is somewhat complicated.

A different approach to obtain some of the bounds in [17] was presented in [8]. The appeal of this new approach is in showing the possibility to work with Delsarte's linear inequalities without resorting to language and tools of orthogonal polynomial theory. In particular, [8] establishes a connection between packing bounds and isoperimetric questions in the Hamming cube. To describe this connection, we need a notion of the *fractional edge boundary size* of a subset of the cube³.

For $A \subseteq \{0, 1\}^n$, the fractional edge boundary size of A is defined as

$$|\partial^* A| = \min \left\{ \mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 : \text{supp}(f) \subseteq A, \mathbb{E} f^2 = \frac{|A|}{2^n} \right\} \quad (4)$$

The right hand side of this definition computes a minimum of the variation of f over a certain family of functions. Note that this family contains the characteristic function of A whose variation equals the (normalized) cardinality of the edge boundary ∂A . Consequently, the fractional boundary of A is at most as large as its boundary.

The following result is proved in [18, 19], following the approach of [8]. Let us mention that this claim was proved in [8] for an important special case of linear codes.

Theorem 1.1: *Let A be a subset of the boolean cube $\{0, 1\}^n$ such that*

$$|\partial^* A| \leq (2d + 1) \cdot \frac{|A|}{2^{n-1}}$$

Let C be a binary error-correcting code with minimal distance d . Then

$$|C| \leq n|A|$$

³This is a reformulation of a closely related notion of the maximal eigenvalue of a subset introduced in [8].

This suggests a way to obtain upper bounds for codes by finding subsets of the cube with a small fractional boundary. Natural candidates to try are the isoperimetric sets, that is Hamming balls and subcubes. Their fractional boundaries are analyzed in [8]. It turns out that among these two options, a Hamming ball has the smaller fractional edge boundary. Note that this pinpoints an intriguing difference between the notions of the fractional edge-boundary and that of the 'ordinary' edge-boundary, for which the subcubes are the optimal sets.

Let B denote a Hamming ball of radius r . It is shown in [8] that

$$|\partial^* B| \leq 4 \left(\frac{n}{2} - \sqrt{r(n-r)} + o(n) \right) \cdot \frac{|B|}{2^n} \quad (5)$$

Combined with Theorem 1.1, this shows that a binary error-correcting code with minimal distance d is at most as large, up to negligible multiplicative factors, as a Hamming ball of radius $r = n/2 - \sqrt{d(n-d)}$. This provides an alternative proof of the *first linear programming bound* for binary codes [17].

This concludes our detour into coding theory. We are now ready to state the isoperimetric problem of [8].

1.3 An isoperimetric problem for the Hamming cube

In order to obtain the best possible bounds on codes via Theorem 1.1, we need to find subsets of the Hamming cube with the smallest possible fractional edge-boundary. In particular, an existence of subsets whose fractional boundary is noticeably smaller than that of Hamming balls of the same cardinality, would imply an improvement on the best currently known bounds. This naturally leads to the following questions [8]

A fractional edge-isoperimetric problem for the Hamming cube:

- What is the smallest possible fractional boundary of a subset of $\{0, 1\}^n$ of a given cardinality?
- Which sets have the smallest fractional boundaries?

These questions were the starting point of our investigation. Before describing our results, let us mention a connection to the classical Faber-Krahn inequality in \mathbb{R}^n , as pointed out in [8].

First, here is a brief description of the Euclidean space inequality, following [5]. For an open set Ω in \mathbb{R}^n , consider the functional

$$F[\phi] = \frac{\|grad \phi\|_2^2}{\|\phi\|_2^2} \quad (6)$$

where ϕ ranges over smooth functions supported in Ω , and the associated infimum $\lambda^*(\Omega) = \inf_{\phi} F[\phi]$.

$\lambda^*(\Omega)$ is referred to as the fundamental tone of Ω . The Faber-Krahn inequality states that among all sets Ω of the same measure, Euclidean ball has the minimal fundamental tone.

In the discrete setting of the Hamming cube, a reasonable interpretation of (6) is to consider the functional

$$F[f] = \frac{\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2}{\mathbb{E} f^2}$$

where f ranges over functions supported in a subset A of $\{0,1\}^n$. In our terminology, the “fundamental tone” of A is

$$\lambda^*(A) = \min_f F[f] = \frac{2^n}{|A|} \cdot |\partial^* A|$$

Hence, the set with the smallest fractional boundary for its size has the minimal fundamental tone, and vice versa.

Following [8] we will refer to the fractional edge-isoperimetric problem as the discrete Faber-Krahn problem for the Hamming cube.

1.4 Main results

Our main technical result is a modified version of the logarithmic Sobolev inequality (2). Let $H(x) = -x \log x - (1-x) \log(1-x)$ be the “natural” (i.e., using natural logarithms) entropy function.

Theorem 1.2:

- Let $f : \{0,1\}^n \rightarrow \mathbb{R}$ be a non-zero function, and let $\rho = \frac{1}{n} \frac{\text{Ent}(f^2)}{\mathbb{E} f^2}$. Then

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq C(\rho) \cdot \text{Ent}(f^2), \quad (7)$$

where

$$C(x) = \frac{4}{x} \cdot \left(\frac{1}{2} - \sqrt{H^{-1}(\log 2 - x) \left(1 - H^{-1}(\log 2 - x) \right)} \right)$$

- The function $C(\cdot)$ is an increasing convex function, taking $[0, \log 2]$ to $[2, 2/\log 2]$.

Inequality (7) is tight in the following sense: for each $\rho \in [0, \log 2]$ there exists a non-constant function $f = f_\rho$ such that $\text{Ent}(f^2) \geq \rho n \mathbb{E} f^2$ and

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \leq (1 + o_n(1)) \cdot C(\rho) \cdot \text{Ent}(f^2) \quad (8)$$

This follows from the tightness of inequality (10) below (see the second part of Theorem 1.4), since that inequality is a corollary of (7). The fact that (10) is tight for Hamming balls follows from (5). The functions f_ρ are the minimal variation functions supported on a Hamming ball of an appropriate radius. They are constructed explicitly in [8].

Theorem 1.2 together with the observation $\text{Ent}(f^2) \geq \mathbb{E} f^2 \log \frac{\mathbb{E} f^2}{\mathbb{E}^2 |f|}$ ([7]), implies a corresponding modification of the functional isoperimetric inequality (3).

Corollary 1.3:

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq C(\rho) \cdot \mathbb{E} f^2 \log, \frac{\mathbb{E} f^2}{\mathbb{E}^2 |f|} \quad (9)$$

where $\rho = \frac{1}{n} \log \frac{\mathbb{E} f^2}{\mathbb{E}^2 |f|}$. Inequality (9) is tight in the same sense and for the same reasons (8) is tight.

Let us briefly discuss this inequality. It implies, in particular, that as the ratio $\frac{\mathbb{E} f^2}{\mathbb{E}^2 f}$ grows (the function f becomes less "flat") its edge-isoperimetric constant approaches the isoperimetric constant $C = \frac{2}{\log 2}$ in the edge-isoperimetric inequality (3) for 0-1 functions. One possible partial explanation for this phenomenon is that, for functions supported on a small set $A \subseteq \{0, 1\}^n$, the main contribution to the variation $\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2$ is likely to come from edges (x, y) which belong to the edge-boundary of A .

It is now straightforward to derive the fractional edge-isoperimetric inequality (10) from Corollary 1.3. Let f be a function supported on a subset A of $\{0, 1\}^n$. Then, by the Cauchy-Schwarz inequality, $\frac{\mathbb{E} f^2}{\mathbb{E}^2 f} \geq \frac{2^n}{|A|}$. Since the function $C(\cdot)$ is monotone, we have

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geq C \left(\frac{1}{n} \log \frac{2^n}{|A|} \right) \cdot \log \frac{2^n}{|A|} \cdot \mathbb{E} f^2$$

Recalling the definition of the fractional edge boundary (4), and substituting the explicit expression for $C(\cdot)$, gives (10).

The inequality (10), together with the second part of Theorem 1.4, provide an asymptotic solution to the Faber-Krahn problem for the Hamming cube, at least in the range of interest to the coding theory. It turns out that, up to an error which becomes negligible as the dimension n grows, Hamming balls of radius $0 \leq r \leq \frac{n}{2} - o(n)$ are the sets with the smallest fractional boundary (fundamental tone) for their size.

From the viewpoint of coding theory, this implies that Theorem 1.1 cannot lead to an improvement on the best currently known bounds for binary codes [17]. Let us briefly discuss one implication of this fact. The bounds in [17] are obtained following Delsarte's linear programming approach, and there are claims in coding theory ([2, 18]) which seem to indicate that these are the best bounds attainable with this approach. Since Theorem 1.1 is derived within the same linear programming framework, inequality (10) may be interpreted as an additional evidence in this direction⁴. It seems worthwhile to point out that, in this manner, the coding theory provides both the question prompting this investigation and an indication of what the answer might be, by suggesting the putative optimality of Hamming balls for the Faber-Krahn problem.

Theorem 1.4:

⁴and it is not, in this sense, very surprising.

- Let $H(x) = -x \log x - (1-x) \log(1-x)$ be the entropy function. Then for any subset A of $\{0,1\}^n$ holds

$$|\partial^* A| \geq 4n \left(\frac{1}{2} - \sqrt{H^{-1} \left(\frac{\log |A|}{n} \right) \left(1 - H^{-1} \left(\frac{\log |A|}{n} \right) \right)} \right) \cdot \frac{|A|}{2^n} \quad (10)$$

- On the other hand, let B be a Hamming ball. Then

$$|\partial^* B| \leq 4n \left(\frac{1}{2} - \sqrt{H^{-1} \left(\frac{\log |B|}{n} \right) \left(1 - H^{-1} \left(\frac{\log |B|}{n} \right) \right)} + o_n(1) \right) \cdot \frac{|B|}{2^n}$$

The second part of this theorem is due to [8]. It follows from (5) and the fact that the cardinality of a Hamming ball of radius r is at least $\exp\{nH(\frac{r}{n}) - o(n)\}$ [16].

The bound in (10) is not very good for balanced subsets A , for which $\frac{\log |A|}{n}$ is close to $\log 2$. For instance, for $|A| = 2^{n-1}$, the bound gives $|\partial^* A| \geq \log 2$. On the other hand, it is not hard to see that the correct bound in this case is $|\partial^* A| \geq 1$. Equality is attained on a subcube of co-dimension one (but not on a Hamming ball of radius $n/2$).

On the other hand, (10) is interesting, as long as the error term $o(1)$ in the second part of Theorem 1.4 is of order lower than that of the main term. This error term is of order $n^{-1/2}$, up to poly-logarithmic terms⁵. Therefore, the theorem provides a satisfactory lower bound for the fractional edge boundary size of A as long as

$$\left| \frac{\log |A|}{n} - \log 2 \right| \geq \tilde{\Omega}(n^{-1/2})$$

In particular, Hamming balls of radius at most $n/2 - \tilde{\Omega}(n^{3/4})$ have (asymptotically) the smallest fractional edge-boundary for their size.

Questions:

- How small can the fractional edge boundary size of a balanced subset A of $\{0,1\}^n$ be?
- Which balanced sets have the smallest fractional edge-boundary?

Finally, let us briefly mention an additional application of inequality (7), of a somewhat different nature. The 'standard' logarithmic Sobolev inequality (2) is used in [9] to derive a hypercontractive inequality for functions on the discrete cube (see also [4, 3]). In order to describe this inequality, let $\{w_S\}_{S \in \{0,1\}^n}$ be the character basis in the space of real-valued functions on the Hamming cube. Let $0 \leq t \leq 1$ and let $T = T_t$ be a linear operator taking a function $f = \sum_S \hat{f}(S) w_S$ to $Tf = \sum_S t^{|S|} \hat{f}(S) w_S$. Then ([9])

$$\|Tf\|_2 \leq \|f\|_{1+t^2} \quad (11)$$

⁵This could be derived from the computation in [8], or, alternatively, from the estimates on the minimal roots of Krawchouk polynomials [7, 18].

Substituting (7) instead of (2) in the proof in [9] leads to a modified version of (11). It turns out that the exponent 2 on the right hand side of the inequality can be replaced by a function $e(\rho)$, depending on $\rho = \frac{1}{n} \frac{\text{Ent}(f^2)}{\mathbb{E}f^2}$. The function $e(\rho)$ is decreasing, with $e(0) = 2$.⁶ This modified inequality might be useful in coding theory, following the applications of (11) in [13, 1].

2 The proof of Theorem 1.2

Let us start with a brief overview. The main goal of this section is to prove the logarithmic Sobolev inequality (7). Our proof follows the outline of the proof of (2) in [9]. We will prove an inequality (13), which will imply (7) as a corollary, first for the base case $n = 1$, and then for general n , using subadditivity of entropy. Compared to [9], we need to prove a bit more for the base case (see Remark 2.2 below) and to carry this additional information along to the general case. This seems to complicate things somewhat, making it necessary to go through the intermediate inequality (13).

We will prove the second claim of the theorem, on the properties of the function $C(\rho)$ in (7), along the way, in Lemma 2.1.

We may and will assume from now on that we deal only with nonnegative functions on $\{0, 1\}^n$, since substituting $|f|$ instead of f in (7) decreases the left hand side and does not affect the right hand side.

Several functions on the real line play an important role in the proof. We start by defining these functions and stating some of their properties. Let ψ be defined on $[0, 1]$ by

$$\psi(t) = \frac{1}{2}(1 - \sqrt{t})^2 \log(1 - \sqrt{t})^2 + \frac{1}{2}(1 + \sqrt{t})^2 \log(1 + \sqrt{t})^2 - (1 + t) \log(1 + t) \quad (12)$$

In other words, $\psi(t) = \text{Ent}(f^2)$, where f is a function on $\{0, 1\}$ with $f(0) = 1 - \sqrt{t}$, $f(1) = 1 + \sqrt{t}$.

The following main technical lemma lists the relevant properties of ψ and several derived functions, including the function C . The proof of the lemma is rather long and is postponed till the next section.

Lemma 2.1:

1. *The function ψ is strictly increasing and concave on $[0, 1]$, taking this interval onto $[0, 2 \log 2]$. This allows us to define the inverse function $\phi = \psi^{-1}$. This is a strictly increasing convex function taking $[0, 2 \log 2]$ onto $[0, 1]$.*
2. *The function $\psi(t)/(1 + t)$ is strictly increasing and concave on $[0, 1]$, taking this interval onto $[0, \log 2]$. This allows us to define the inverse function $\alpha(t) = \left(\frac{\psi(t)}{1+t}\right)^{-1}$. This is a strictly increasing convex function taking $[0, \log 2]$ to $[0, 1]$.*

⁶Hence, for functions with high entropy, this gives a strengthening of (11).

3. The function $c(t) = \frac{4\alpha(t)}{t(1+\alpha(t))}$ is strictly increasing and convex on $[0, \log 2]$, taking this interval onto $[2, 2/\log 2]$.⁷

4. The function $c(t)$ has an explicit representation

$$c(t) = \frac{4}{t} \cdot \left(\frac{1}{2} - \sqrt{H^{-1}(\log 2 - t) \left(1 - H^{-1}(\log 2 - t) \right)} \right).$$

In other words $c = C$, where C is the function in (7).

Note that the second claim of Theorem 1.2 follows from the third and the fourth claims of this lemma.

Remark 2.2: We mentioned a difference between the proof of the base case $n = 1$ here and in [9]. Let us give some details. In [9], (2) is shown for $\{0, 1\}$, which, in our notation, amounts to proving an inequality $\psi(t) \leq 2t$ on $[0, 1]$. We show, in addition, that ψ is concave on $[0, 1]$. This additional convexity property turns out to be crucial for our proof. ■

Next, we introduce additional notation, and prove a simple auxiliary inequality.

For a function f on $\{0, 1\}^n$ let $D^2(f) = \mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2$, and let $K^2(f) = \frac{1}{4} \cdot \mathbb{E}_x \sum_{y \sim x} (f(x) + f(y))^2$. Note that $K^2(f) = n\mathbb{E}f^2 - \frac{1}{4} \cdot D^2(f)$. Note also that for a non-zero nonnegative function f , $K^2(f)$ is strictly positive.

Lemma 2.3: For a nonnegative function f on $\{0, 1\}^n$ holds

$$\text{Ent}(f^2) \leq 2 \log 2 \cdot K^2(f)$$

Proof: We will prove the claim for the base case $n = 1$ and then use subadditivity of entropy to deduce it for the general case.

The case $n = 1$. We may assume f is non-zero, since the claim is trivially true otherwise. Both sides of the inequality are 2-homogeneous, and consequently we may assume $\mathbb{E}f = 1$. Without loss of generality, $f(0) = 1 - s$, $f(1) = 1 + s$ for some $0 \leq s \leq 1$. We need to show

$$\text{Ent}(f^2) = \psi(s^2) \leq 2 \log 2 \cdot K^2(f) = 2 \log 2,$$

and this is indeed true by the first claim of Lemma 2.1, since ψ is increasing and $\psi(1) = 2 \log 2$.

The general case $n \geq 1$. Let $1 \leq i \leq n$ be an index of a coordinate and let $x \in \{0, 1\}^n$. Fixing all the coordinates $j \neq i$ to be x_j we obtain a copy of $\{0, 1\}$. Let $f_i^{(x)}$ be the restriction of f to this one-dimensional cube.

Recall ([14]) that entropy is subadditive, namely

$$\sum_{i=1}^n \mathbb{E}_x \text{Ent}(f_i^{(x)}) \geq \text{Ent}(f),$$

⁷Here, as usual, we take $c(0) = \lim_{t \rightarrow 0} \frac{4\alpha(t)}{t(1+\alpha(t))} = 2$.

while $D^2(f)$ is additive, that is

$$\sum_{i=1}^n \mathbb{E}_x D^2 \left(f_i^{(x)} \right) = D^2(f).$$

$K^2(f)$ is also additive, since $\sum_{i=1}^n \mathbb{E}_x K^2 \left(f_i^{(x)} \right) = \sum_{i=1}^n \mathbb{E}_x \left(\mathbb{E} \left(f_i^{(x)} \right)^2 - 1/4 D^2 \left(f_i^{(x)} \right) \right) = n \mathbb{E} f^2 - 1/4 D^2(f) = K^2(f)$.

From this and the base case, we have

$$Ent(f^2) \leq \sum_{i=1}^n \mathbb{E}_x Ent \left(\left(f_i^{(x)} \right)^2 \right) \leq 2 \log 2 \cdot \sum_{i=1}^n \mathbb{E}_x K^2 \left(f_i^{(x)} \right) = 2 \log 2 \cdot K^2(f)$$

■

Let us now pass to the main technical claim in this section. Note that the right hand side in (13) is well-defined, due to the preceding lemma.

Proposition 2.4: *Let f be a non-zero nonnegative function on $\{0, 1\}^n$. Then*

$$D^2(f) \geq 4 \cdot K^2(f) \phi \left(\frac{Ent(f^2)}{K^2(f)} \right) \quad (13)$$

The proof follows the same outline. First we prove the base case $n = 1$. In this case, we will show equality

$$D^2(f) = 4 \cdot K^2(f) \phi \left(\frac{Ent(f^2)}{K^2(f)} \right).$$

Indeed, since f is non-zero, we may assume $\mathbb{E}f = 1$. This implies $K^2(f) = 1$. Thus we have to prove $D^2(f) = 4\phi(Ent(f^2))$. Let $0 \leq s \leq 1$, and $f(0) = 1 - s$, $f(1) = 1 + s$. Then $D^2(f) = 4s^2$ and $4\phi(Ent(f^2)) = 4\phi(\psi(s^2)) = 4s^2$, verifying the base case.

We also need to deal with a slight technicality, the case f is the zero function, because in the general case below, some of the one-dimensional restrictions of f might be zero. In this case, we formally define $Ent(f^2)/K^2(f)$ to be zero. Then (13) remains valid (as equality) in the one-dimensional case. It is easy to see that this formal definition does not affect the computation below.

The general case. Let $n \geq 1$. Then, by the base case, by subadditivity of the entropy, and by convexity and monotonicity of ϕ :

$$D^2(f) = \sum_{i=1}^n \mathbb{E}_x D^2 \left(f_i^{(x)} \right) = 4 \cdot \sum_{i=1}^n \mathbb{E}_x K^2 \left(f_i^{(x)} \right) \phi \left(\frac{Ent \left(\left(f_i^{(x)} \right)^2 \right)}{K^2 \left(f_i^{(x)} \right)} \right) =$$

$$4K^2(f) \cdot \sum_{i=1}^n \mathbb{E}_x \frac{K^2(f_i^{(x)})}{K^2(f)} \phi \left(\frac{Ent \left((f_i^{(x)})^2 \right)}{K^2(f_i^{(x)})} \right) \geq$$

$$4K^2(f) \cdot \phi \left(\frac{1}{K^2(f)} \cdot \sum_{i=1}^n \mathbb{E}_x Ent \left((f_i^{(x)})^2 \right) \right) \geq 4K^2(f) \cdot \phi \left(\frac{Ent(f^2)}{K^2(f)} \right)$$

■

We proceed to derive (7) from (13). By homogeneity, we may assume $\mathbb{E}f^2 = 1$. Let $0 \leq \rho \leq \log 2$. Consider the functional

$$R[f] = \frac{D^2(f)}{Ent(f^2)}$$

where f ranges over the non-empty⁸ compact set of nonnegative non-zero functions satisfying $\mathbb{E}f^2 = 1$ and $Ent(f^2) \geq \rho n$, and the associated minimum

$$m(\rho) = \min_f R[f]$$

To complete the proof of (7) and of Theorem 1.2, we will show

$$m(\rho) \geq c(\rho), \tag{14}$$

where c is the function defined in the third part of Lemma 2.1.

Indeed, fix ρ and let $m = m(\rho)$. Let f be a function at which $R[f]$ attains its minimum, that is: $\mathbb{E}f^2 = 1$, $Ent(f^2) \geq \rho n$ and $D^2(f) = m Ent(f^2)$. Then

$$D^2(f) = m Ent(f^2) \geq m \rho n.$$

This means $K^2(f) = n \mathbb{E}f^2 - 1/4 D^2(f) \leq (1 - (m\rho)/4) n$, and therefore $\frac{Ent(f^2)}{K^2(f)} \geq \frac{4\rho}{4 - m\rho}$.

Recall ϕ is an increasing convex function on $[0, 2 \log 2]$ with $\phi(0) = 0$. Therefore the function $\tau(y) = \phi(y)/y$ is increasing. This implies

$$\phi \left(\frac{Ent(f^2)}{K^2(f)} \right) = \frac{Ent(f^2)}{K^2(f)} \tau \left(\frac{Ent(f^2)}{K^2(f)} \right) \geq \frac{Ent(f^2)}{K^2(f)} \tau \left(\frac{4\rho}{4 - m\rho} \right)$$

and, by Proposition 2.4,

$$D^2(f) \geq 4K^2(f) \phi \left(\frac{Ent(f^2)}{K^2(f)} \right) \geq 4\tau \left(\frac{4\rho}{4 - m\rho} \right) Ent(f^2)$$

which means

$$m \geq 4\tau \left(\frac{4\rho}{4 - m\rho} \right)$$

⁸Recall $Ent(f^2) \geq \mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{\mathbb{E}^2 f}$ ([7]).

The rest is simple algebra. Recall $\tau = \phi(y)/y$ and $\psi = \phi^{-1}$. Since ψ is increasing, the last inequality is equivalent to

$$\psi\left(\frac{m\rho}{4-m\rho}\right) \geq \frac{4\rho}{4-m\rho}.$$

Substituting $y = \frac{m\rho}{4-m\rho}$, this translates to $\psi(y)/(y+1) \geq \rho$. Recall $\alpha = \left(\frac{\psi(y)}{1+y}\right)^{-1}$. Since α is increasing, we obtain

$$\frac{m\rho}{4-m\rho} \geq \alpha(\rho)$$

This is the same as

$$m \geq \frac{4\alpha(\rho)}{\rho(\alpha(\rho) + 1)}.$$

This is equivalent to (14), completing the proof of (7) and of Theorem 1.2. ■

3 Proof of Lemma 2.1

Let

$$h(t) = \frac{1}{2}(1-t)^2 \log(1-t)^2 + \frac{1}{2}(1+t)^2 \log(1+t)^2 - (1+t^2) \log(1+t^2). \quad (15)$$

In other words, $h(t) = \psi(t^2)$. We will start with some useful properties of the function h .

Lemma 3.1:

1.

$$h' \geq h \geq 0$$

2.

$$(1-t^2) h' \geq t h''$$

Proof: We have $h'(t) = 2 \cdot ((1+t) \log(1+t) - (1-t) \log(1-t) - t \log(1+t^2))$ and $h''(t) = \frac{4}{1+t^2} - 2 \log\left(\frac{1+t^2}{1-t^2}\right)$.

The first claim of the lemma is easy. Nonnegativity of h follows from nonnegativity of ψ , and

$$h'(t) - h(t) = (1-t^2) \log(1+t) - (1-t)(3-t) \log(1-t) + (1-t^2) \log(1+t^2) \geq 0$$

The second claim is somewhat harder. We have, rearranging and simplifying:

$$\frac{1}{2} \cdot ((1-t^2) h' - t h'') = t^3 \log\left(\frac{1+t^2}{1-t^2}\right) + (1-t^2) \log\left(\frac{1+t}{1-t}\right) - \frac{2t}{1+t^2}$$

That is, we need to show

$$(1+t^2)t^3 \log\left(\frac{1+t^2}{1-t^2}\right) + (1-t^4) \log\left(\frac{1+t}{1-t}\right) \geq 2t$$

In fact, an even stronger inequality

$$t^3 \log\left(\frac{1+t^2}{1-t^2}\right) + (1-t^4) \log\left(\frac{1+t}{1-t}\right) \geq 2t$$

is valid for $t \in [0, 1]$. This is easy to check for $t = 1$. To see this for $0 \leq t < 1$, recall that for $-1 < x < 1$ holds $\log\left(\frac{1+x}{1-x}\right) = 2 \sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1}$. Substituting these series in the inequality above, we need to show

$$t^3 \cdot \sum_{k=0}^{\infty} \frac{t^{4k+2}}{2k+1} + (1-t^4) \cdot \sum_{k=0}^{\infty} \frac{t^{2k+1}}{2k+1} \geq t,$$

and this is easily verified by observing that all the higher coefficients of the power series on the left hand side are nonnegative. ■

We pass to the proof of Lemma 2.1

Claim 1

We will prove ψ is concave by showing ψ'' is negative on $(0, 1)$. We have for $0 < t < 1$:

$$\psi''(t) = -\frac{1}{4t^{3/2}} \cdot h'(\sqrt{t}) + \frac{1}{4t} \cdot h''(\sqrt{t}) < 0$$

The last inequality follows from the second claim of Lemma 3.1.

To see that ψ is increasing, it suffices to verify $\psi' \geq 0$ at 1. Indeed, $\psi'(1) = \frac{1}{2}h'(1) = \log 2 > 0$, completing the proof of Claim 1.

Claim 2

Let $\xi(t) = \frac{\psi}{1+t}$. We will verify $\xi' > 0$ on $(0, 1)$. Indeed,

$$\xi' = \frac{(1+t)\psi' - \psi}{(1+t)^2}$$

so we need to check that $\psi' > \frac{\psi}{1+t}$. Substituting $\psi(t) = h(\sqrt{t})$, this amounts to checking

$$h'(\sqrt{t}) \geq \frac{2\sqrt{t}}{1+t} \cdot h(\sqrt{t})$$

Since $\frac{2\sqrt{t}}{1+t} < 1$ on $(0, 1)$, it suffices to show $h' \geq h$, which is true by the first claim of Lemma 3.1.

To show concavity of ξ , we will prove that

$$-\xi'' > 2\xi, \tag{16}$$

implying $\xi'' < 0$ on $(0, 1)$. Direct calculation gives

$$\xi''(t) = \frac{(1+t)^2\psi'' - 2((1+t)\psi' - \psi)}{(1+t)^3}$$

Simplifying, $-\xi'' > 2\xi'$ reduces to

$$2t\psi > 2t(1+t)\psi' + (1+t)^2\psi''$$

Since $\psi'' < 0$ we have $(1+t)^2\psi'' < 4t\psi''$. Therefore, it suffices to prove

$$\psi \geq (1+t)\psi' + 2\psi''$$

Again, since ψ is concave with $\psi(0) = 0$, we have $\psi \geq t\psi'$. Therefore, we only need to prove

$$-2\psi'' \geq \psi'$$

Writing this in terms of h , this is equivalent to $(1-s^2)h' \geq sh''$ which is given by the second claim of Lemma 3.1.

Claim 4

We will show this claim before the third claim of the lemma. That claim is somewhat more involved, and its proof is relegated to the end of this section.

Here we need to verify

$$\frac{\alpha(t)}{1+\alpha(t)} = \frac{1}{2} - \sqrt{H^{-1}(\log 2 - t)(1 - H^{-1}(\log 2 - t))}$$

for all $t \in [0, \log 2]$. This is equivalent to

$$H^{-1}(\log 2 - t) = \frac{1}{2} - \sqrt{\frac{\alpha}{1+\alpha} \left(1 - \frac{\alpha}{1+\alpha}\right)} = \frac{(1 - \sqrt{\alpha})^2}{2(1+\alpha)}$$

Recall that $\alpha = \alpha(t)$ is defined to satisfy $t = \frac{\psi(\alpha)}{1+\alpha}$, where $\psi(\alpha) = Ent(f^2)$, and f is a function on $\{0, 1\}$ with $g(0) = 1 - \sqrt{\alpha}$, $g(1) = 1 + \sqrt{\alpha}$. It is not hard to verify the identity

$$\frac{\psi(\alpha)}{1+\alpha} = \ln 2 - H \left(\frac{(1 - \sqrt{\alpha})^2}{2(1+\alpha)} \right)$$

for all $\alpha \in [0, 1]$, and we are done.

Claim 3

First, we show that c is increasing. Direct computation gives that c' is positive on $(0, \log 2)$ iff $t\alpha' > \alpha + \alpha^2$ on this interval. Both sides of this inequality are 0 at 0, and we compare derivatives, that is, show $t\alpha'' > 2\alpha\alpha'$.

Since α is convex with $\alpha(0) = 0$, we have $\alpha \leq t\alpha'$ in the interval. Hence, it suffices to show $\alpha'' > 2(\alpha')^2$.

Recall $\alpha = \xi^{-1}$. Consequently, $\alpha'(\xi(t)) = \frac{1}{\xi'(t)}$, and $\alpha''(\xi(t)) = -\frac{\xi''(t)}{(\xi'(t))^3}$. Therefore, $\alpha'' > 2(\alpha')^2$ is equivalent to $-\xi'' > 2\xi'$, which is given by (16).

It remains to show that c is convex. This turns out to be significantly harder than the other proofs in this Section. We provide a somewhat sketchy argument below.

Direct computation shows that $c'' > 0$ on $(0, \log 2)$ iff

$$t^2(1 + \alpha)\alpha'' + 2\alpha(1 + \alpha)^2 > 2t^2(\alpha')^2 + 2t(1 + \alpha)\alpha' \quad (17)$$

First, we rewrite this inequality in terms of $\xi = \alpha^{-1}$. Let $t = \xi(x)$, that is $\alpha(t) = x$, $\alpha'(t) = \frac{1}{\xi'(x)}$, $\alpha''(t) = -\frac{\xi''(x)}{(\xi'(x))^3}$. Substituting and simplifying, one gets

$$-(1 + x)\xi^2\xi'' + 2x(1 + x)^2(\xi')^3 > 2\xi^2\xi' + 2(1 + x)\xi(\xi')^2,$$

which has to hold for all x in $(0, 1)$.

Next, we rewrite this in terms of $\psi = (1 + x)\xi$, obtaining

$$(1 + x)\psi^2(-\psi'') > 2((1 + x)\psi' - \psi)^2(\psi - x\psi')$$

Note that all the expressions in the brackets are positive, since ψ is concave and ξ is increasing, as we saw in the proofs of Claims 1 and 2 above. We simplify this inequality, replacing ψ with $x\psi'$ on the left hand side and in the first term on the right hand side, and arriving to the stronger inequality

$$x(1 + x)\psi(-\psi'') > 2\psi'(\psi - x\psi')$$

We rewrite this in terms of the function h , defined in (15) above. As in the proof of Claim 1, expressing ψ and its derivatives in terms of h , leads to the following equivalent inequality:

$$2x(h')^2 > (3 - x^2)hh' + x(1 + x^2)hh'' \quad (18)$$

From now on we concentrate on the proof of (18). It will be convenient to write h and its derivatives in terms of two new functions $L_1(x) = \log \frac{1+x}{1-x}$ and $L_2(x) = \log \frac{1+x^2}{1-x^2}$. Recalling the expressions for h and its derivatives (as in the proof of Lemma 3.1 above, we have

- $h(x) = 2xL_1 - (1 + x^2)L_2$
- $h'(x) = 2L_1 - 2xL_2$
- $h''(x) = \frac{4}{1+x^2} - 2L_2$

Rewriting (18) in terms of L_1 and L_2 , and simplifying, one arrives to

$$(3 - x^2)(1 + x^2)L_1L_2 + 2x(1 + x^2)L_2 > 2x(1 - x^2)L_1^2 + 4xL_2^2 + 4x^2L_1$$

We expand both sides of this inequality as power series for $x \in (0, 1)$. Recall that $L_1(x) = 2\sum_{k=0}^{\infty} \frac{1+x^{2k+1}}{2k+1}$, and, consequently, $L_2(x) = 2\sum_{k=0}^{\infty} \frac{1+x^{4k+2}}{2k+1}$. Therefore, both sides of this inequality have only odd terms.

Let the left hand side be equal to

$$F(x) = 4 \cdot \sum_{k=0}^{\infty} \ell_{2k+1} x^{2k+1}$$

and the right hand side be equal to

$$G(x) = 4 \cdot \sum_{k=0}^{\infty} r_{2k+1} x^{2k+1}$$

We will argue that

1. All the coefficients ℓ_{2k+1} and r_{2k+1} are nonnegative.
2. $\ell_1 = r_1 = 0$, $\ell_3 = r_3 = \ell_5 = r_5 = 4$.
3. For all odd k starting from $k = 3$:

$$\ell_{2k+1} > r_{2k+1} \quad \text{and} \quad \ell_{2k+1} + \ell_{2k+3} > r_{2k+1} + r_{2k+3}$$

This will imply

$$\begin{aligned} F(x) - G(x) &= 4 \cdot \sum_{k=3}^{\infty} (\ell_{2k+1} - r_{2k+1}) x^{2k+1} = 4 \cdot \sum_{\text{odd } k \geq 3} ((\ell_{2k+1} - r_{2k+2}) - (\ell_{2k+3} - r_{2k+3}) x^2) \cdot x^{2k+1} > \\ &4 \cdot \sum_{\text{odd } k \geq 3} (\ell_{2k+1} - r_{2k+2}) (1 - x^2) \cdot x^{2k+1} > 0, \end{aligned}$$

completing the proof of (18) and of Claim 3. Hence it remains to prove the properties of the coefficients.

In fact, the coefficients can be computed explicitly, which makes it possible to verify the required properties. We omit the (easy but cumbersome) details. For completeness sake, we do list explicit expressions for the coefficients below⁹.

- For an odd $k \geq 3$:

$$\begin{aligned} \ell_{2k+1} &= \left(\frac{8k-20}{(2k-3)(2k+1)} \right) \cdot \sum_{m=1}^{(k-1)/2} \frac{1}{4m-3} + \frac{4}{2k-1} \cdot \sum_{m=1}^{k-2} \frac{1}{2m+1} + \\ &\left(\frac{3}{2k+1} + \frac{2}{2k-1} - \frac{1}{2k-3} \right) \cdot \sum_{m=1}^{(k-1)/2} \frac{1}{2m-1} + \left(\frac{1}{k} + \frac{3}{k(2k+1)} + \frac{6}{(2k-1)(2k+1)} \right) \end{aligned}$$

and

$$r_{2k+1} = \frac{2k+2}{k(2k-1)} - \frac{2}{k(k-1)} \cdot \sum_{m=1}^{k-1} \frac{1}{2m-1}$$

⁹Our apologies to the reader.

- For an even $k \geq 4$:

$$\begin{aligned} \ell_{2k+1} = & \left(\frac{8k-20}{(2k-3)(2k+1)} \right) \cdot \sum_{m=1}^{k-2} \frac{1}{2m+1} + \frac{4}{2k-1} \cdot \sum_{m=1}^{k/2} \frac{1}{4m-3} + \\ & \left(\frac{3}{2k+1} + \frac{2}{2k-1} - \frac{1}{2k-3} \right) \cdot \sum_{m=1}^{(k-2)/2} \frac{1}{2m-1} + \left(\frac{1}{k-1} + \frac{6}{(2k-1)(2k+1)} + \frac{10k-1}{(k-1)(2k-1)(2k+1)} \right) \\ \text{and} \\ r_{2k+1} = & \frac{8}{k} \cdot \sum_{m=1}^{k/2} \frac{1}{2m-1} + \frac{2k+2}{k(2k-1)} - \frac{2}{k(k-1)} \cdot \sum_{m=1}^{k-1} \frac{1}{2m-1} \end{aligned}$$

It remains to compute the coefficients for $k = 1, 2$. This is easily done directly, verifying the property 2 above.

References

- [1] A. Ashikhmin, G. Cohen, M. Krivelevich, S. Litsyn, *Bounds on distance distributions in codes of known size*, IEEE Trans. Inform. Theory, vol. IT-51 2005, 250-258.
- [2] A. Barg, D. B. Jaffe, *Numerical results on the asymptotic rate of binary codes*, in **Codes and Association schemes**, (A. Barg and S. Litsyn, eds.), Amer. Math. Soc., Providence 2001.
- [3] W. Beckner, *Inequalities in Fourier Analysis*, Annals of Math., 102(1975), pp. 159-182.
- [4] A. Bonami, *Etude des coefficients Fourier des fonctiones de $L^p(G)$* , Ann. Inst. Fourier (Grenoble) 20:2 (1970), pp. 335-402.
- [5] I. Chavel, **Isoperimetric inequalities**, Cambridge University Press, 2001.
- [6] P. Delsarte, *An algebraic approach to association schemes of coding theory*, Philips Res. Rep. Suppl. 10, 1973.
- [7] D. Falik, A. Samorodnitsky, *A combinatorial proof of a theorem of Kahn, Kalai, and Linial*, CPC, to appear.
- [8] J. Friedman and J-P. Tillich, *Generalized Alon-Boppana Theorems and Error-Correcting Codes*, preprint, 2002.
- [9] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. of Math., 97, 1975, pp. 1061-1083.
- [10] L. H. Harper, *Optimal numberings and isoperimetric problems on graphs*, J. Combin. Theory, 1, 1966, pp. 385-393.
- [11] L. H. Harper, *Optimal assignment of numbers to vertices*, J. Soc. Ind. Appl. Math., 12, 1964, pp. 131-135.

- [12] S. Hart, *A note on the edges of the n -cube*, Discr. Math., 14, 1976, pp. 157-163.
- [13] G. Kalai, N. Linial, *On the distance distribution of codes*, IEEE Trans. Inform. Theory, vol. IT-41 1995, 1467-1472 (see, in particular, the extended version of the paper on <http://www.cs.huji.ac.il/~nati/>).
- [14] M. Ledoux, *Concentration of measure and logarithmic Sobolev inequalities*, Seminaire de Probabilites, XXXIII, Lecture Notes in Math., 1709, Springer 1990, pp. 120-216.
- [15] V. I. Levenshtein, *Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces*, IEEE Trans. Inform. Theory, vol. IT-41, 1995, 1303-1321.
- [16] J.H. van Lint, **Introduction to Coding Theory**, Third edition. Springer-Verlag, Berlin, 1999.
- [17] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory, vol. IT-23, 1977, 157-166.
- [18] M. Navon, A. Samorodnitsky, *On Delsarte's linear programming bounds for binary codes*, Proceedings of FOCS 46.
- [19] M. Navon, A. Samorodnitsky, *Linear programming bounds via a covering argument*, Disc. and Comp. Geometry, to appear.
- [20] G. Szegő, **Orthogonal Polynomials**, American Mathematical Society, 1939.